18.175 Theory of Probability
Fall 2008

Section 2

Random variables and their properties. Expectation.

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $(\mathcal{S}, \mathcal{B})$ be a measurable space where $\mathcal{B}$ is a σ-algebra of subsets of $\mathcal{S}$. A *random variable* $X : \Omega \to \mathcal{S}$ is a measurable function, i.e.

$$B \in \mathcal{B} \Longrightarrow X^{-1}(B) \in \mathcal{A}.$$

When $\mathcal{S} = \mathbb{R}$ we will usually consider a σ-algebra $\mathcal{B}$ of Borel measurable sets generated by sets $\bigcup_{i \le n}(a_i, b_i]$ (or, equivalently, generated by sets (a_i, b_i) or by open sets).

Lemma 3 *$X : \Omega \to \mathbb{R}$ is a random variable iff for all $t \in \mathbb{R}$*

$$\{X \le t\} := \{\omega \in \Omega : X(\omega) \in (-\infty, t]\} \in \mathcal{A}.$$

Proof. Only $\Leftarrow$ direction requires proof. We will prove that

$$\mathcal{D} = \{D \subseteq \mathbb{R} : X^{-1}(D) \in \mathcal{A}\}$$

is a σ-algebra. Since sets $(-\infty, t] \in \mathcal{D}$ this will imply that $\mathcal{B} \subseteq \mathcal{D}$. The result follows simply because taking pre-image preserves set operations. For example, if we consider a sequence $D_i \in \mathcal{D}$ for $i \ge 1$ then

$$X^{-1}\Bigl(\bigcup_{i\ge 1} D_i\Bigr) = \bigcup_{i \ge 1} X^{-1}(D_i) \in \mathcal{A}$$

because $X^{-1}(D_i) \in \mathcal{A}$ and $\mathcal{A}$ is a σ-algebra. Therefore, $\bigcup_{i\ge 1} D_i \in \mathcal{D}$. Other properties can be checked similarly, so $\mathcal{D}$ is a σ-algebra.

$\square$

Let us define a measure $\mathbb{P}_X$ on $\mathcal{B}$ by $\mathbb{P}_X = \mathbb{P} \circ X^{-1}$, i.e. for $B \in \mathcal{B}$,

$$\mathbb{P}_X(B) = \mathbb{P}(X \in B) = \mathbb{P}(X^{-1}(B)) = \mathbb{P} \circ X^{-1}(B).$$

$(\mathcal{S}, \mathcal{B}, \mathbb{P}_X)$ is called the *sample space* of a random variable X and $\mathbb{P}_X$ is called *the law* of X. Clearly, on this space a random variable $\xi : \mathcal{S} \to \mathcal{S}$ defined by the identity $\xi(s) = s$ has the same law as X.

When $\mathcal{S} = \mathbb{R}$, a function $F(t) = \mathbb{P}(X \le t)$ is called the cumulative distribution function (c.d.f.) of X.

Lemma 4 *F is a c.d.f. of some r.v. X iff*

1. $0 \le F(t) \le 1$,
2. *F is non-decreasing, right-continuous,*

3. $\lim_{t\to-\infty} F(t) = 0$, $\lim_{t\to+\infty} F(t) = 1$.

Proof. The fact that any c.d.f. satisfies properties 1 - 3 is obvious. Let us show that F which satisfies properties 1 - 3 is a c.d.f. of some r.v. X. Consider algebra A consisting of sets $\bigcup_{i\le n}(a_i, b_i]$ for disjoint intervals and for all $n \ge 1$. Let us define a function $\mathbb{P}$ on A by

$$\mathbb{P}\Big(\bigcup_{i\le n}(a_i, b_i]\Big) = \sum_{i\le n}\big(F(a_i) - F(b_i)\big).$$

One can show that $\mathbb{P}$ is countably additive on A. Then, by Caratheodory extension Theorem 1, $\mathbb{P}$ extends uniquely to a measure $\mathbb{P}$ on $\sigma(A) = \mathcal{B}$ - Borel measurable sets. This means that $(\mathbb{R}, \mathcal{B}, \mathbb{P})$ is a probability space and, clearly, random variable $X : \mathbb{R} \to \mathbb{R}$ defined by $X(x) = x$ has c.d.f. $\mathbb{P}(X \le t) = F(t)$. Below we will sometimes abuse the notations and let F denote both c.d.f. and probability measure $\mathbb{P}$.
Alternative proof. Consider a probability space $([0,1], \mathcal{B}, \lambda)$, where λ is the Lebesgue measure. Define r.v. $X : [0,1] \to \mathbb{R}$ by the quantile transformation

$$X(t) = \inf\{x \in \mathbb{R}, F(x) \ge t\}.$$

The c.d.f. of X is $\lambda(t : X(t) \le a) = F(a)$ since

$$X(t) \le a \Longleftrightarrow \inf\{x : F(x) \ge t\} \le a \Longleftrightarrow \exists a_n \to a, F(a_n) \ge t \Longleftrightarrow F(a) \ge t.$$

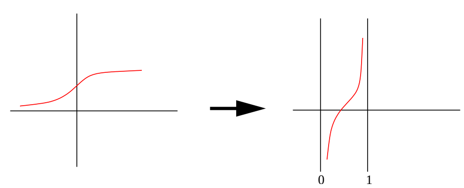

Figure 2.1: A random variable defined by quantile transformation.

□

Definition. Given a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and a r.v. $X : \Omega \to \mathcal{S}$ let $\sigma(X)$ be a σ-algebra generated by a collection of sets $\{X^{-1}(B) : B \in \mathcal{B}\}$. Clearly, $\sigma(X) \subseteq \mathcal{A}$. Moreover, the above collection of sets is itself a σ-algebra. Indeed, consider a sequence $A_i = X^{-1}(B_i)$ for some $B_i \in \mathcal{B}$. Then

$$\bigcup_{i\ge1} A_i = \bigcup_{i\ge1} X^{-1}(B_i) = X^{-1}\Big(\bigcup_{i\ge1} B_i\Big) = X^{-1}(B)$$

where $B \in \bigcup_{i\ge1} B_i \in \mathcal{B}$. $\sigma(X)$ is called the *σ-algebra generated by a r.v. X*.

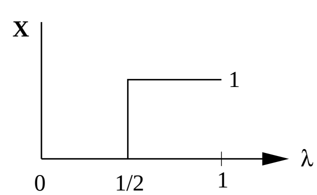

Figure 2.2: $\sigma(X)$ generated by X.

Example. Consider a r.v. defined in figure 2.2. We have $\mathbb{P}(X = 0) = \frac{1}{2}, \mathbb{P}(X = 1) = \frac{1}{2}$ and

$$\sigma(X) = \Big\{\emptyset, \Big[0, \frac{1}{2}\Big], \Big(\frac{1}{2}, 1\Big], [0,1]\Big\}.$$

Lemma 5 *Consider a probability space* $(\Omega, \mathcal{A}, \mathbb{P})$*, a measurable space* $(\mathcal{S}, \mathcal{B})$ *and random variables* $X : \Omega \to \mathcal{S}$ *and* $Y : \Omega \to \mathbb{R}$*. Then the following are equivalent:*

1. $Y = g(X)$ *for some (Borel) measurable function* $g : \mathcal{S} \to \mathbb{R}$*.*
2. $Y : \Omega \to \mathbb{R}$ *is measurable on* $(\Omega, \sigma(X))$*, i.e. with respect to the* σ*-algebra generated by* X*.*

Remark. It should be obvious from the proof that $\mathbb{R}$ can be replaced by any separable metric space.

Proof. The fact that 1 implies 2 is obvious since for any Borel set $B \subseteq \mathbb{R}$ the set $B' := g^{-1}(B) \in \mathcal{B}$ and, therefore,

$$\{Y = g(X) \in B\} = \{X \in g^{-1}(B) = B'\} = X^{-1}(B') \in \sigma(X).$$

Let us show that 2 implies 1. For all integer n and k consider sets

$$A_{n,k} = \Big\{\omega : Y(\omega) \in \Big[\frac{k}{2^n}, \frac{k+1}{2^n}\Big)\Big\} = Y^{-1}\Big(\Big[\frac{k}{2^n}, \frac{k+1}{2^n}\Big)\Big).$$

By 2, $A_{n,k} \in \sigma(X) = \{X^{-1}(B) : B \in \mathcal{B}\}$ and, therefore, $A_{n,k} = X^{-1}(B_{n,k})$ for some $B_{n,k} \in \mathcal{B}$. Let us consider a function

$$g_n(X) = \sum_{k\in\mathbb{Z}} \frac{k}{2^n}\ \mathrm{I}(X \in B_{n,k}).$$

By construction, $|Y - g_n(X)| \le \frac{1}{2^n}$ since

$$Y(\omega) \in \Big[\frac{k}{2^n}, \frac{k+1}{2^n}\Big) \Longleftrightarrow X(\omega) \in B_{n,k} \Longleftrightarrow g_n(X(\omega)) = \frac{k}{2^n}.$$

It is easy to see that $g_n(x) \le g_{n+1}(x)$ and, therefore, $g(x) = \lim_{n\to\infty} g_n(x)$ is a measurable function on $(\mathcal{S}, \mathcal{B})$ and, clearly, $Y = g(X)$.

□

Discrete random variables.

A r.v. $X : \Omega \to \mathcal{S}$ is called discrete if $\mathbb{P}_X(\{S_i\}_{i\ge1}) = 1$ for some sequence $S_i \in \mathcal{S}$.

□

Absolutely continuous random variables.

On a measure space $(\mathcal{S}, \mathcal{B})$, a measure $\mathbb{P}$ is called *absolutely continuous* w.r.t. a measure λ if

$$\forall B \in \mathcal{B}, \lambda(B) = 0 \Longrightarrow \mathbb{P}(B) = 0.$$

The following is a well known result from measure theory.

Theorem 2 *(Radon-Nikodym) If* $\mathbb{P}$ *and* λ *are sigma-finite and* $\mathbb{P}$ *is absolutely continuous w.r.t.* λ *then there exists a* Radon-Nikodym derivative $f \ge 0$ *such that for all* $B \in \mathcal{B}$

$$\mathbb{P}(B) = \int_B f(s)d\lambda(s).$$

f *is uniquely defined up to a* λ*-null sets.*

In a typical setting of $\mathcal{S} = \mathbb{R}^k$, a probability measure $\mathbb{P}$ and Lebesgue's measure λ, f is called the *density* of the distribution $\mathbb{P}$.

□

Independence.

Consider a probability space $(\Omega, \mathcal{C}, \mathbb{P})$ and two σ-algebras $\mathcal{A}, \mathcal{B} \subseteq \mathcal{C}$. $\mathcal{A}$ and $\mathcal{B}$ are called *independent* if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) \ \text{ for all } \ A \in \mathcal{A}, B \in \mathcal{B}.$$

σ-algebras $\mathcal{A}_i \subseteq \mathcal{C}$ for $i \leq n$ are *independent* if

$$\mathbb{P}(A_1 \cap \dots \cap A_n) = \prod_{i \leq n} \mathbb{P}(A_i) \ \text{ for all } \ A_i \in \mathcal{A}_i.$$

σ-algebras $\mathcal{A}_i \subseteq \mathcal{C}$ for $i \leq n$ are *pairwise independent* if

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j) \ \text{ for all } \ A_i \in \mathcal{A}_i, A_j \in \mathcal{A}_j, i \neq j.$$

Random variables $X_i : \Omega \to \mathcal{S}$ for $i \leq n$ are (pairwise) independent if σ-algebras $\sigma(X_i), i \leq n$ are (pairwise) independent which is just another convenient way to state the familiar

$$\mathbb{P}(X_1 \in B_1, \dots, X_n \in B_n) = \mathbb{P}(X_1 \in B_1) \times \dots \times \mathbb{P}(X_n \in B_n)$$

for any events $B_1, \dots, B_n \in \mathcal{B}$.

Example. Consider a regular tetrahedron die, Figure 2.3, with red, green and blue sides and a red-green-blue base. If we roll this die then indicators of different colors provide an example of pairwise independent r.v.s that are not independent since

$$\mathbb{P}(r) = \mathbb{P}(b) = \mathbb{P}(g) = \frac{1}{2} \text{ and } \mathbb{P}(rb) = \mathbb{P}(rg) = \mathbb{P}(bg) = \frac{1}{4}$$

but

$$\mathbb{P}(rbg) = \frac{1}{4} \neq \mathbb{P}(r)\mathbb{P}(b)\mathbb{P}(g) = \Big(\frac{1}{2}\Big)^3.$$

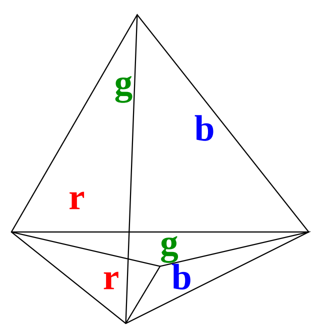

Figure 2.3: Pairwise independent but not independent r.v.s.

□

Independence of σ-algebras can be checked on generating algebras:

Lemma 6 *If algebras $A_i, i \leq n$ are independent then σ-algebras $\sigma(A_i)$ are independent.*

Proof. Obvious by Approximation Lemma 2.

□

Lemma 7 *Consider r.v.s $X_i : \Omega \to \mathbb{R}$ on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$.*

1. *X_i's are independent iff*

$$\mathbb{P}(X_1 \leq t_1, \dots, X_n \leq t_n) = \mathbb{P}(X_1 \leq t_1) \times \dots \times \mathbb{P}(X_n \leq t_n). \tag{2.0.1}$$

2. *If the laws of X_i's have densities $f_i(x)$ then X_i's are independent iff a joint density exists and*

$$f(x_1, ..., x_n) = \prod f_i(x_i).$$

Proof. 1 is obvious by Lemma 6 because (2.0.1) implies the same equality for intervals

$$\mathbb{P}(X_1 \in (a_1, b_1], \dots, X_n \in (a_n, b_n]) = \mathbb{P}(X_1 \in (a_1, b_1]) \times \dots \times \mathbb{P}(X_n \in (a_n, b_n])$$

and, therefore, for finite union of disjoint such intervals. To check this for intervals (for example, for $n = 2$) we can write $\mathbb{P}(a_1 < X_1 \le b_1, a_2 < X_n \le b_2)$ as

$$\begin{aligned} & \mathbb{P}(X_1 \le b_1, X_2 \le b_2) - \mathbb{P}(X_1 \le a_1, X_2 \le b_2) - \mathbb{P}(X_1 \le b_1, X_2 \le a_2) + \mathbb{P}(X_1 \le a_1, X_2 \le a_2) \\ = \; & \mathbb{P}(X_1 \le b_1)\mathbb{P}(X_2 \le b_2) - \mathbb{P}(X_1 \le a_1)\mathbb{P}(X_2 \le b_2) - \mathbb{P}(X_1 \le b_1)\mathbb{P}(X_2 \le a_2) + \mathbb{P}(X_1 \le a_1)\mathbb{P}(X_2 \le a_2) \\ = \; & \big(\mathbb{P}(X_1 \le b_1) - \mathbb{P}(X_1 \le a_1)\big)\big(\mathbb{P}(X_2 \le b_2) - \mathbb{P}(X_2 \le a_2)\big) = \mathbb{P}(a_1 < X_1 \le b_1)\mathbb{P}(a_2 < X_2 \le b_2). \end{aligned}$$

To prove 2 we start with "$\Longleftarrow$".

$$\begin{aligned} \mathbb{P}(\cap\{X_i \in A_i\}) &= \mathbb{P}(\boldsymbol{X} \in A_1 \times \cdots \times A_n) = \int_{A_1 \times \cdots \times A_n} \prod f_i(x_i) d\boldsymbol{x} \\ &= \prod \int_{A_i} f_i(x_i) dx_i \text{ \{by Fubini's Theorem\}} = \prod_{i \le n} \mathbb{P}(X_i \in A_i). \end{aligned}$$

Next, we prove "$\Longrightarrow$". First of all, by independence,

$$\mathbb{P}(\boldsymbol{X} \in A_1 \times \cdots \times A_n) = \prod \mathbb{P}(X_i \in A_i) \overset{\text{Fubini}}{=} \int_{A_1 \times \cdots \times A_n} \prod f_i(x_i) d\boldsymbol{x}.$$

Therefore, the same equality holds for sets in algebra A that consists of finite unions of disjoint sets $A_1 \times \cdots \times A_n$, i.e.

$$\mathbb{P}(\boldsymbol{X} \in B) = \int_B \prod f_i(x_i) d\boldsymbol{x} \text{ for } B \in A.$$

Both $\mathbb{P}(X \in B), \int_B \prod f_i(x_i) d\boldsymbol{x}$ are countably additive on A and finite,

$$\mathbb{P}(\mathbb{R}^n) = \int_{\mathbb{R}^n} \prod f_i(x_i) d\boldsymbol{x} = 1.$$

By the Caratheodory extension Theorem 1, they extend uniquely to all Borel sets $\mathcal{B} = \sigma(A)$, so

$$\mathbb{P}(B) = \int_B \prod f_i(x_i) d\boldsymbol{x} \text{ for } B \in \mathcal{B}.$$

□

Expectation. If $X : \Omega \to \mathbb{R}$ is a random variable on $(\Omega, \mathcal{A}, \mathbb{P})$ then *expectation* of X is defined as

$$\mathbb{E}X = \int_\Omega X(\omega) d\mathbb{P}(\omega).$$

In other words, expectation is just another term for the integral with respect to a probability measure and, as a result, expectation has all the usual properties of the integrals. Let us emphasize some of them.

Lemma 8 *1. If F is the c.d.f. of X then for any measurable function $g : \mathbb{R} \to \mathbb{R}$,*

$$\mathbb{E}g(x) = \int_\mathbb{R} g(x) dF(x).$$

2. If X is discrete, i.e. $\mathbb{P}(X \in \{x_i\}_{i \ge 1}) = 1$, then

$$\mathbb{E}X = \sum_{i \ge 1} x_i \mathbb{P}(X = x_i).$$

3. *If $X : \Omega \to \mathbb{R}^k$ has a density $f(x)$ on $\mathbb{R}^k$ and $g : \mathbb{R}^k \to \mathbb{R}$ then*

$$\mathbb{E}g(X) = \int g(x)f(x)dx.$$

Proof. All these properties follow by making a change of variables $x = X(\omega)$ or $\omega = X^{-1}(x)$, i.e.

$$\mathbb{E}g(X) = \int_{\Omega} g(X(\omega))d\mathbb{P}(\omega) = \int_{\mathbb{R}} g(x)d\mathbb{P} \circ X^{-1}(x) = \int_{\mathbb{R}} g(x)d\mathbb{P}_X(x),$$

where $\mathbb{P}_X = \mathbb{P} \circ X^{-1}$ is the law of X. Another way to see this would be to start with indicator functions of sets $g(x) = \mathrm{I}(x \in B)$ for which

$$\mathbb{E}g(X) = \mathbb{P}(X \in B) = \mathbb{P}_X(B) = \int_{\mathbb{R}} \mathrm{I}(x \in B)d\mathbb{P}_X(x)$$

and, therefore, the same is true for simple step functions

$$g(x) = \sum_{i \geq n} w_i \mathrm{I}(x \in B_i)$$

for disjoint B_i. By approximation, this is true for any measurable functions.

$\square$